INTERNATIONAL SEARCH REPORT

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 G06F17/30

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, WPI Data, PAJ, INSPEC, COMPENDEX, IBM-TDB

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category°	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DEWDNEY N ET AL: "The Form is the Substance: Classification of Genres in Text" ACL 2001 CONFERENCE: WORKSHOP ON HUMAN LANGUAGE TECHNOLOGY AND KNOWLEDGE MANAGEMENT, 'Online! 6 - 7 July 2001, XP002280032 Toulouse, France Retrieved from the Internet: <URL:http://www.elsnet.org/km2001/dewdney. pdf> 'retrieved on 2004-05-12! abstract section 1.2, paragraphs 2, 3 section 1.3, paragraph 4 section 2.1, paragraphs 1, 2 section 3, paragaph 1 section 3.2, paragraph 1 section 4, paragaph 2; table 3 ___ -/--	1-14

| X | Further documents are listed in the continuation of box C. | | X | Patent family members are listed in annex. |

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
17 May 2004	01/06/2004

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Stauch, M

ternational Application No

'CT/GB 20/0400014

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	MC CALLUM A ET AL: "A Comparison of Event Models for Naive Bayes Text Classification" AAAI 1998: FIFTEENTH NATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE: WORKSHOP ON LEARNING FOR TEXT CATEGORIZATION (W7), 'Online! 27 July 1998 (1998-07-27), XP002280033 Madison, Wisconsin, USA Retrieved from the Internet: <URL:http://www.cs.umass.edu/{mccallum/papers/multinomial-aaai98w.ps> 'retrieved on 2004-05-12! abstract "Introduction", paragraphs 1-6 section "Probabilistic Framework of Navie Bayes", paragraph 2 section "Multivariate Bernoulli Model"	1-14
X	SEBASTANI F: "Machine Learning in Automated Text Categorization" AVM COMPUTING SURVEYS, 'Online! vol. 34, no. 1, March 2002 (2002-03), pages 1-47, XP002280034 Retrieved from the Internet: <URL:http://portal.acm.org/ft_gateway.cfm?id=505283&type=pdf&coll=GUIDE&dl=ACM&CFID=21427216&CFTOKEN=89202948> 'retrieved on 2004-05-12! abstract page 7, right-hand column, line 1 - line 14 page 8, right-hand column, line 31 -page 9, left-hand column, line 4 page 9, left-hand column, line 35 - line 42 page 10, right-hand column, line 19 - line 39 page 11, right-hand column, line 7 -page 12, left-hand column, line 24 page 18, right-hand column, line 19 - line 40 page 30, left-hand column, line 1 - line 17	1-14
X	WO 00 26795 A (JUSTSYSTEM PITTSBURGH RESEARCH) 11 May 2000 (2000-05-11) abstract page 3, line 23 - line 31 page 4, line 28 -page 5, line 9 page 5, line 31 -page 6, line 11 page 6, line 26 - line 30 page 10, line 20 -page 11, line 2	1-14

-/--

3

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	GLOVER E: "Using Extra-Topical User Preferences to improve Web-based Metasearch" ONLINE DISSERTATION, 'Online! 2001, XP002280518 University of Michigan Retrieved from the Internet: <URL:http://www.webir.org/resources/phd/Glover_2001.pdf> 'retrieved on 2004-05-17! abstract -----	1-14

3

Patent document cited in search report		Publication date	Patent family member(s)		Publication date
WO 0026795	A	11-05-2000	AU	1122100 A	22-05-2000
			WO	0026795 A1	11-05-2000